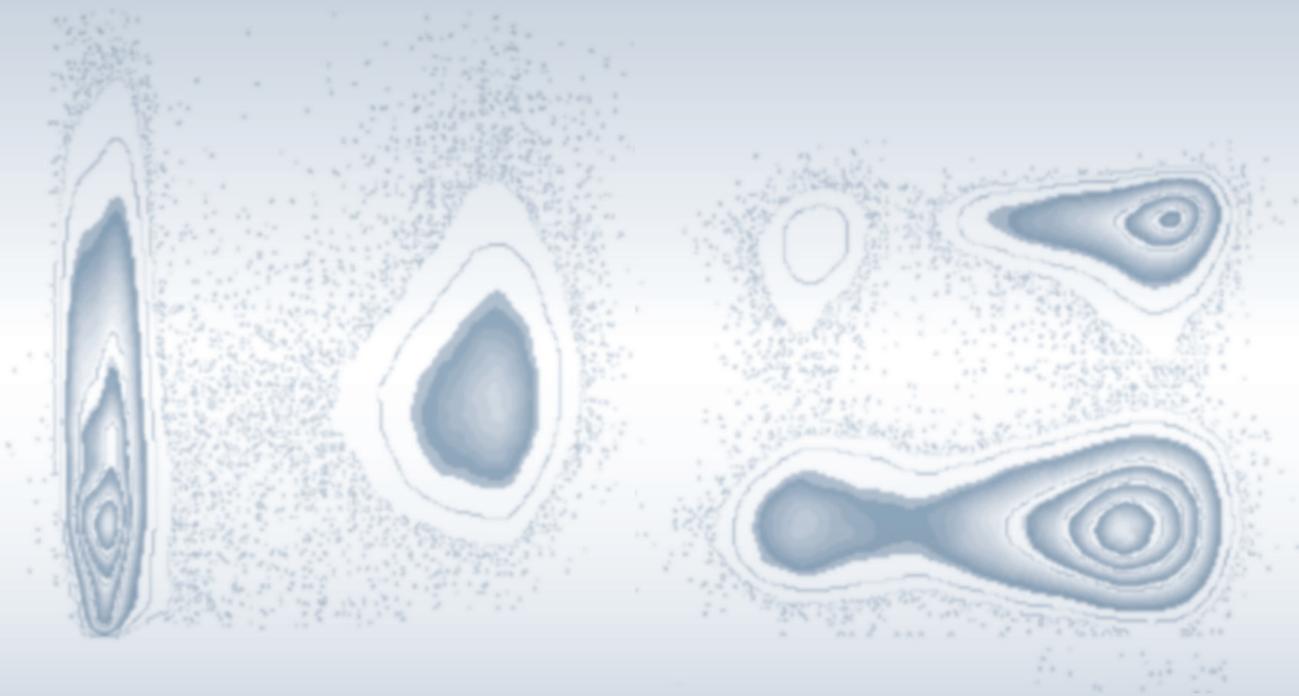


LA CYTOMÉTRIE EN FLUX

2^e édition

Coordonné par

Jean-François Mayol
Corinne Laplace-Builhé
Christian D. Muller



Les données de cytométrie en flux : affichage, contenu, représentation et optimisation

Hélène Dujardin, Corinne Laplace-Builhé

1. Affichage des données de cytométrie en flux
 - 1.1. Visualisation des résultats
 - 1.2. Échelles utilisées
2. Enregistrement des données de cytométrie en flux
 - 2.1. Informations contenues dans les fichiers de cytométrie
 - 2.2. Standards des fichiers de cytométrie
3. outils pour l'analyse supervisée manuelle et optimisation des données de cytométrie
 - 3.1. Bases de la sélection des populations d'intérêt
 - 3.2. Sélection sur les paramètres FSC et SSC et temps : une option préliminaire utile
 - 3.3. Stratégie de gestion des échantillons et des données pour optimiser les résultats
4. Conclusion

Comprendre la signification des données de cytométrie, contrôler la relevance des signaux obtenus, conduite à tenir pour les analyses complexes : voici quelques points clés importants que se propose d'aborder ce chapitre, afin de vous aider à optimiser les informations issues de vos analyses en cytométrie en flux.

1. AFFICHAGE DES DONNÉES DE CYTOMÉTRIE EN FLUX

1.1. Visualisation des résultats

L'affichage graphique des résultats est généralement réalisé grâce au logiciel du cytomètre lui-même ou *via* un logiciel tiers compatible. Les événements sont affichés la plupart du temps sous forme de graphiques mono- ou bi-paramétriques. Dans le cas des graphiques mono-paramétriques ou histogrammes, l'axe des abscisses représente l'intensité relative de fluorescence pour le paramètre analysé, l'axe des ordonnées le nombre d'événements associés à chaque valeur de fluorescence (figure 5-1A).

Lors de l'affichage des données sous forme graphique, le logiciel d'analyse procède à une nouvelle répartition des données en canaux (appelé *binning* en anglais), à

une résolution plus faible (généralement autour de 1 024 canaux) que celle utilisée par le cytomètre lors de la numérisation des données, de façon à améliorer la visualisation des résultats (figure 5-1B). L'histogramme permet à l'expérimentateur d'observer la répartition des événements pour un paramètre donné, de mettre en valeur une distribution bimodale, une distribution en loi normale, ou une distribution asymétrique. En superposant deux histogrammes (*overlay*, en anglais), on peut également mettre en valeur le décalage d'une distribution entre deux échantillons ou deux populations. Dans le cas de l'affichage biparamétrique, l'expérimentateur a accès à la description de deux paramètres de fluorescence simultanément et peut établir des corrélations dans l'expression de ces deux paramètres, ce qui en fait une représentation beaucoup plus instructive dans le cadre d'une expérience multiparamétrique. Cette représentation est souvent privilégiée en cytométrie en flux car elle est plus adaptée à la visualisation de populations de faible fréquence et elle peut aussi permettre de moduler le seuil de négativité du premier paramètre en fonction de l'expression d'un deuxième paramètre. Il existe plusieurs types de graphiques bi-paramétriques, la figure 5-2 montre quelques exemples avec les mêmes données. Le graphique en nuage de point peut afficher l'ensemble des événements individuellement, mais ne permet pas d'apprécier la densité relative des popula-

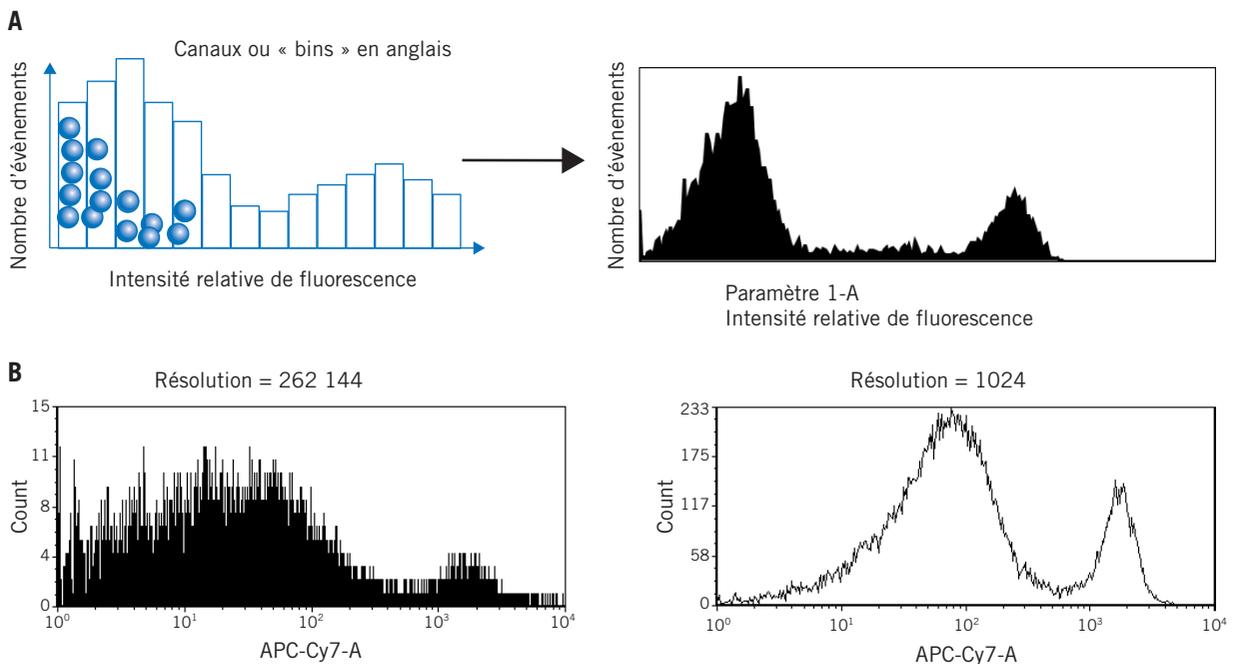


Figure 5-1. Affichage mono-paramétrique des analyses en cytométrie. Il traduit la distribution des événements en fonction d'un paramètre donné. **A** et **B** correspondent à des représentations utilisant des résolutions d'affichage différentes.

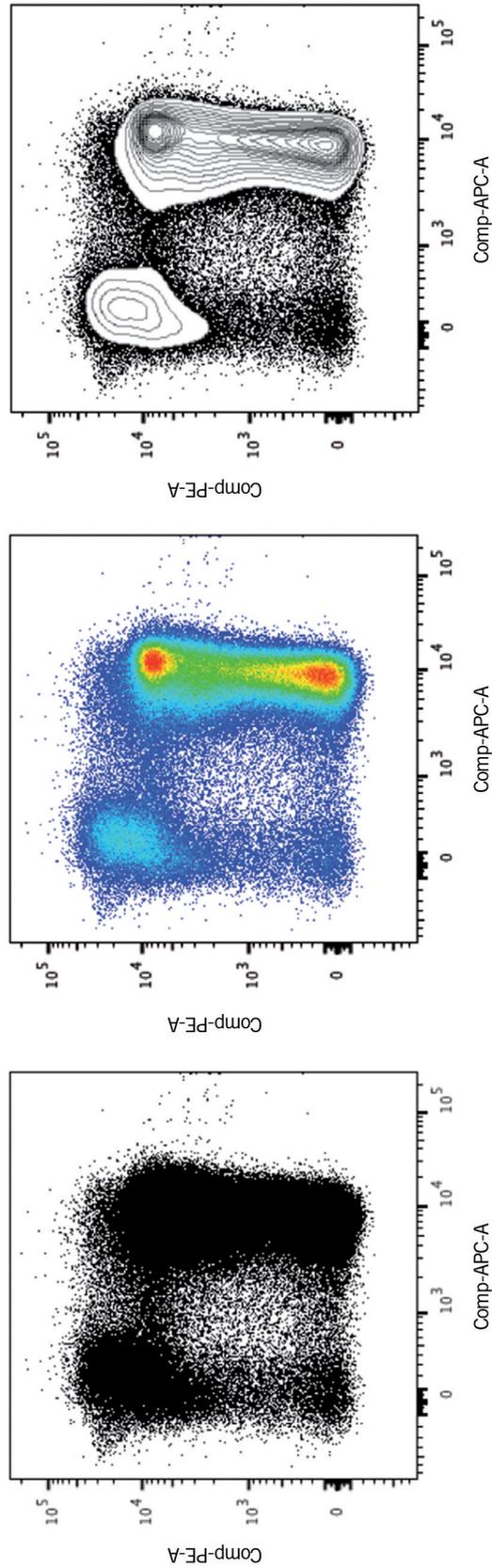


Figure 5-2. Affichage bi-paramétrique des analyses en cytométrie. Il décrit la distribution de deux paramètres observés simultanément (fluorescence versus fluorescence ou fluorescence versus taille par exemple) et permet d'établir des corrélations dans l'expression de ces deux paramètres. Le mode de représentation en nuage de points, densité de couleur ou en contours (de gauche à droite) sera choisi par l'opérateur en fonction de la ou des populations recherchées.

tions dans les zones de forte fréquence, ce qui oblige souvent à réduire la quantité d'événements affichés et donc de réduire l'information disponible. D'autres représentations maintiennent la précision de l'affichage des événements individuels dans les zones de faibles fréquences et indiquent la densité dans les zones de forte fréquence. C'est le cas des graphiques en densité de couleur ou en contours, la densité étant indiquée par la modulation de la couleur pour le premier, par des lignes de contour tracées selon la fréquence des événements pour le deuxième. Des visualisations montrant plus de deux paramètres sont également proposées par les logiciels d'analyse, tels les graphiques en trois dimensions ou même multivariés. Pour les graphiques multiparamétriques, il s'agit alors principalement d'apprécier une image globale d'expression de différents paramètres, un pattern d'expression, et de détecter des variations d'un échantillon à l'autre.

1.2. Échelles utilisées

Plusieurs échelles d'affichages sont utilisées pour afficher les données de cytométrie.

L'échelle linéaire est utilisée lorsque la gamme du signal est relativement faible, comme pour le *scatter* (complexité interne relative) ou pour la mesure de la quantité d'ADN (acide désoxyribonucléique).

Lorsqu'on attend une gamme de signal large, comme pour l'expression antigénique cellulaire, on utilise un affichage en échelle logarithmique, généralement de 4 ou 5 décades. L'échelle logarithmique a aussi l'avantage d'afficher de façon symétrique les distributions log-normales que l'on observe fréquemment dans les systèmes biologiques. Les données peuvent être enregistrées dans le domaine logarithmique dans le fichier de cytométrie après transformation par le cytomètre ou enregistrées dans le domaine linéaire, c'est alors le logiciel qui réalise la transformation logarithmique.

Des alternatives à l'échelle logarithmiques ont été proposées au début des années 2000. L'un des moteurs de ces changements a été l'augmentation du nombre de paramètres analysés et, par conséquent, l'augmentation des chevauchements des signaux dans les canaux de détection non spécifiques. L'affichage de ces résultats en échelle logarithmique posait deux problèmes majeurs : une visualisation médiocre des événements de faible intensité de fluorescence, souvent compressés dans les premiers canaux d'affichage ou, dit de manière plus intuitive, « collés à l'axe », et une impression de sous-compensation pour des données par ailleurs parfaitement compensées (figure 5-3).

Augmenter le nombre de décades d'affichage de l'échelle logarithmique ne permet pas de corriger ces deux problèmes, comme on peut le voir sur la figure 5-3A. Augmenter le voltage des photomultiplicateurs (PMT) au moment de l'acquisition ne fait généralement qu'augmenter le bruit de fond et étaler le signal négatif (non illustré ici). Il fallait donc pouvoir mieux visualiser la partie basse de l'échelle d'affichage, et cela d'autant plus que l'évolution des cytomètres et des fichiers associés a conduit à l'apparition de valeurs négatives d'intensité relative de fluorescence. Ces valeurs négatives sont la conséquence de la soustraction du bruit de fond par certains cytomètres et, dans une plus large mesure, de la compensation du signal sur les valeurs numérisées (compensation digitale).

Plusieurs échelles ont été proposées comme alternative à l'échelle logarithmique, que l'on peut appeler échelles hybrides (exemple d'échelles hybrides : hyperlog, logicle, bi-exponentielle). Ces échelles hybrides permettent d'afficher des valeurs négatives, sont linéaires autour de zéro et redeviennent logarithmiques en s'éloignant du zéro. La figure 5-3C montre un affichage en échelle bi-exponentielle. Grâce à cet affichage, on peut observer aisément l'intégralité des événements et ainsi mieux identifier les populations. Cet affichage permet également de contrôler visuellement la compensation en observant l'alignement des populations et éviter ainsi la tentation de surcompensation. Il est souvent possible de changer les paramètres réglant l'affichage des données en échelle hybride. Ces paramètres influent sur le point de transition entre la partie linéaire et la partie logarithmique de l'échelle, le nombre de décades positives et/ou négatives.

Cette flexibilité permet d'appliquer une échelle adaptée à l'expression de chacun des paramètres de fluorescence étudiés. Cependant, elle peut présenter une difficulté car il n'est pas toujours facile de comprendre la façon dont ces réglages influencent les échelles d'affichage des données, ou même de savoir qu'il est possible de faire ces modifications. En particulier, il est souvent mal compris que l'affichage des données puisse être différent entre le logiciel d'acquisition et les logiciels d'analyse, car ces derniers peuvent utiliser un algorithme différent pour la définition de l'échelle bi-exponentielle ou encore des réglages par défaut différents. La figure 5-3 C à F montre les mêmes données avec des réglages différents de paramètres d'échelle bi-exponentielle. *Les réglages des échelles hybrides doivent donc être manipulés avec précaution.* Ces images mettent également en avant la nécessité d'enregistrer des contrôles appropriés qui permettront de clairement identifier les populations négatives (et donc les populations positives ou intermédiaires).

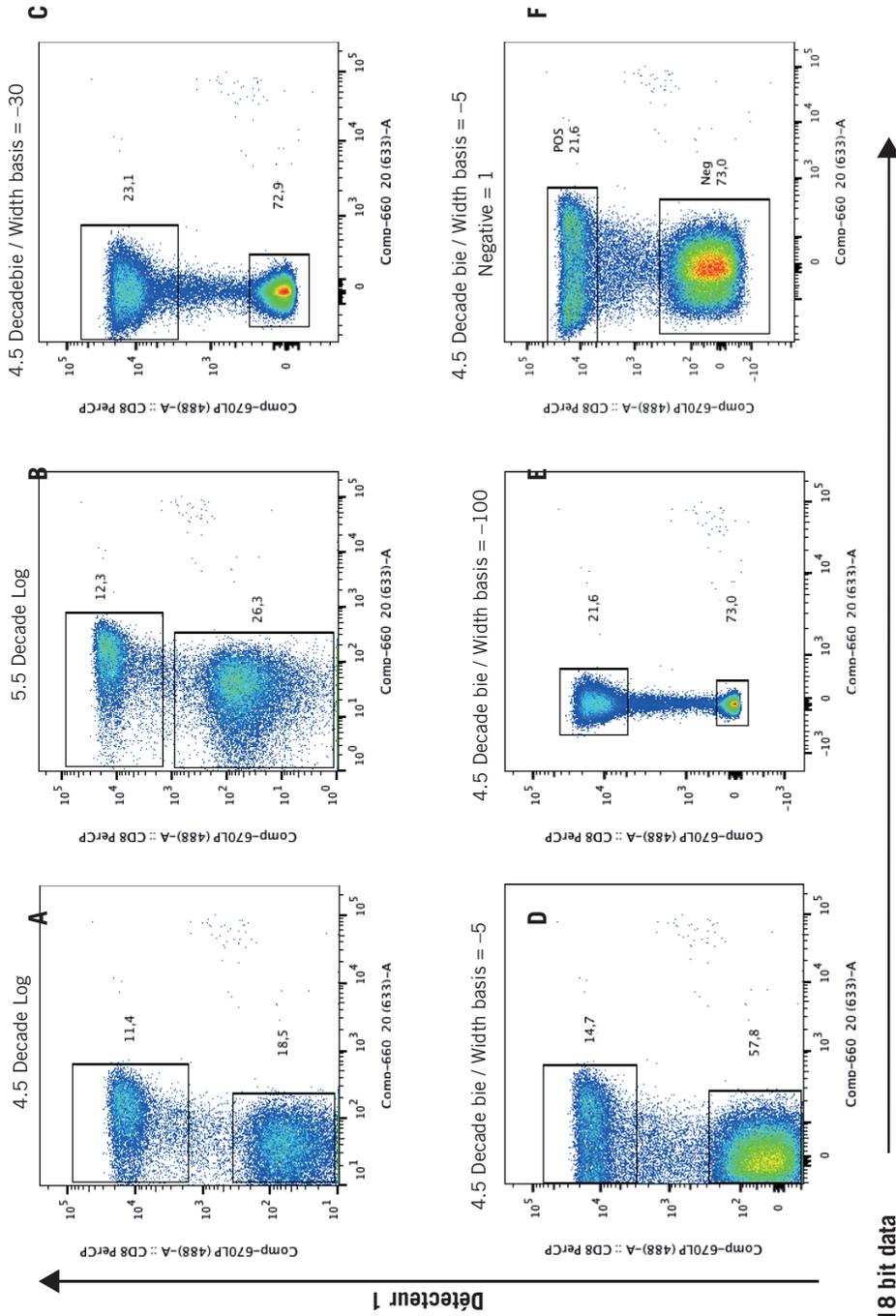


Figure 5-3. Les différentes échelles d'affichage utilisées pour les données multiparamétriques. **(A et B)** Populations marquées avec l'ensemble des fluorochromes de l'expérience sauf le fluorochrome attribué au canal de détection 2. En traçant une région rectangulaire sur les événements bien visualisés du graphique, on observe seulement x % des événements. Une grande proportion des événements n'est donc pas facilement visible par l'expérimentateur. De plus, la population exprimant le plus fort niveau du paramètre 2 semble exprimer également un certain niveau du paramètre 1. Cette impression d'étalement du signal, appelé *spread* en anglais, est un des effets visuels de la compensation. Cette visualisation en échelle logarithmique n'est pas incorrecte, mais elle peut conduire à sous-évaluer les populations négatives en réalisant des régions qui n'incluent pas correctement les premiers canaux d'affichage. Il peut aussi être tentant pour l'expérimentateur de vouloir aligner visuellement les populations P2+ et P2-, ce qui aurait pour conséquence de surcompenser ces données. **(C à F)** C montre un réglage de bi-exponentiel qui permet d'afficher l'ensemble des résultats et de conserver un aspect gaussien aux populations. Les réglages utilisés en **D** ne sont pas suffisants pour afficher l'ensemble des résultats, pour « décoller » les populations de l'axe ; au contraire, en **E**, les populations sont bien décollées mais elles deviennent très compressées, ce qui peut diminuer la qualité de visualisation des populations d'expression intermédiaire. **F** montre un réglage qui permet d'afficher l'ensemble des événements mais qui donne l'impression à l'observateur de l'existence de plusieurs niveaux d'expression. Aucune de ces représentations n'est fautive, mais seul **C** donne un affichage intuitif de la réalité de l'expérience.

2. ENREGISTREMENT DES DONNÉES DE CYTOMÉTRIE EN FLUX

2.1. Informations contenues dans les fichiers de cytométrie

2.1.1. Valeurs enregistrées

Pour chaque événement (cellule, particule, etc.) détecté par le cytomètre, c'est l'ensemble des valeurs numériques des différents paramètres mesurés qui est enregistré et associé à cet événement spécifique (figure 5-4A). Les résultats sont conservés dans un fichier en mode liste (*list mode data file* en anglais). C'est ce qui permet l'analyse simultanée de plusieurs paramètres au niveau unicellulaire et la mise en évidence de corrélations dans l'expression de différents paramètres.

Pour chaque mesure de signal (fluorescence, diffraction de la lumière, etc.), le signal analogique brut obtenu est converti en valeur numérique discrète pour son enregistrement dans le fichier final. La résolution de cette conver-

sion, c'est-à-dire le nombre de canaux utilisés pour décrire le signal analogique, peut varier de 10 bits (1 024 canaux) à 24 bits (16 777 216 canaux) selon les cytomètres. Les valeurs numériques sont enregistrées dans le domaine logarithmique ou linéaire. La mesure choisie est la hauteur (H, *height* en anglais), la largeur (W, *width* en anglais) ou l'aire (A, *area* en anglais) du pic du signal analogique (figure 5-4B). On retrouve l'indication du type de mesure choisie dans le nom du paramètre : P1-H, P1-W ou P1-A. La mesure de l'aire du signal est généralement celle qui est choisie pour l'analyse des paramètres de fluorescence, car elle est plus précise. On peut alors conserver les valeurs H et W pour la discrimination de doublets (classiquement pour les paramètres *forward* et *side scatter*, en anglais). Il est important de garder à l'esprit que les valeurs enregistrées dans le fichier correspondent à des *intensités de fluorescence relatives*. Les valeurs obtenues dépendent de la quantité de lumière mais également des réglages du cytomètre. Seule la mise en place d'outils de calibration permet de parler de valeurs absolues et de comparer directement les intensités de fluorescence obtenues par différents cytomètres. En plus

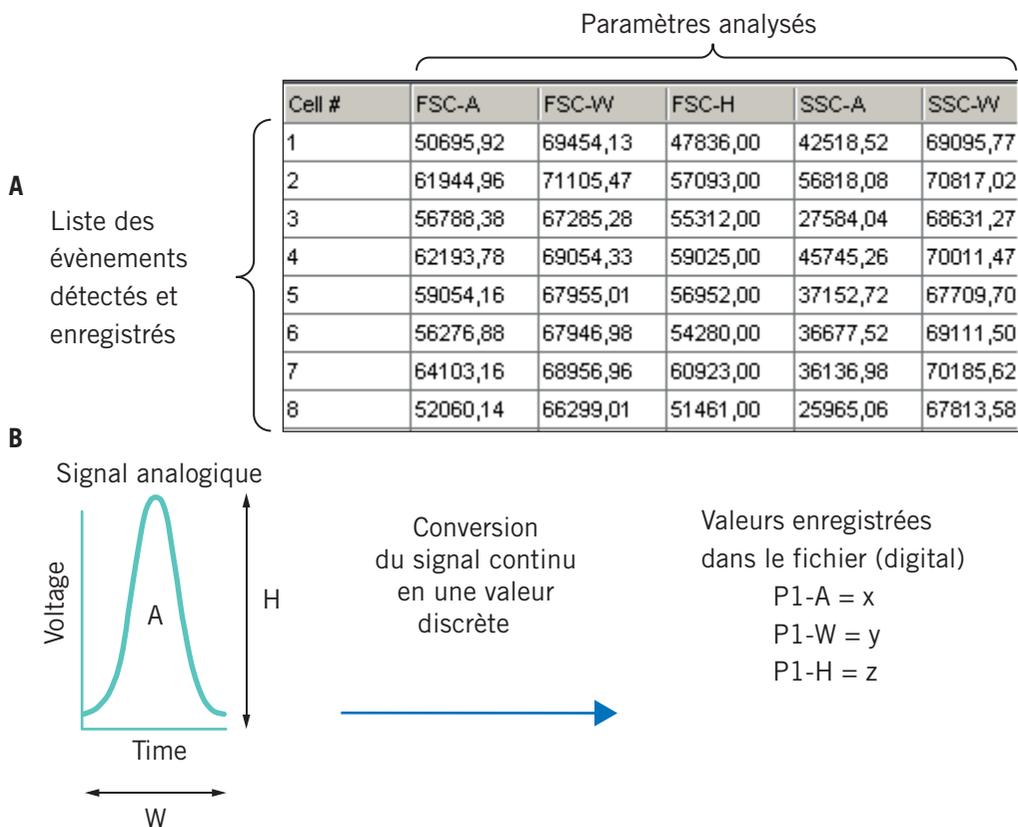


Figure 5-4. Organisation des données enregistrées. **(A)** Structuration et classement des données dans un fichier « *List Mode* » pour chaque événement analysé. **(B)** Paramètres spécifiques pouvant être enregistrés pour chaque signal recueilli (fluorescence ou *scatter*) : H (*height*) hauteur, W (*width*) largeur, ou encore A (*area*) aire, du pic du signal analogique.

des paramètres de fluorescence, la plupart des cytomètres proposent aujourd'hui d'enregistrer le paramètre temps associé aux événements enregistrés ; certains permettent également la mesure du volume de la solution enregistrée ou même l'enregistrement d'images (voir chapitre 27).

2.1.2. Enregistrement de la compensation

L'utilisation simultanée de signaux provenant de plusieurs fluorochromes, lors d'expérience de cytométrie multicouleur par exemple, implique souvent la mise en place de la compensation du signal. L'émission de photons d'un fluorochrome donne lieu à une mesure dans son canal de détection spécifique mais peut également être mesurée par le canal de détection spécifique d'un autre fluorochrome analysé. On parle de chevauchement spectral (*spectral overlap* en anglais), et la méthode pour corriger ce signal non spécifique est appelée compensation. La compensation peut être réalisée avant la numérisation des données, on parle alors de compensation analogique. Elle peut également être réalisée sur les données déjà numérisées, on parle alors de compensation digitale. Dans le cas de la compensation analogique, les valeurs enregistrées dans le fichier de cytométrie sont les valeurs après compensation. Il n'est plus possible d'avoir accès aux valeurs brutes avant compensation. Dans le cas de la compensation digitale, proposée par la majorité des cytomètres actuels, il est possible d'enregistrer dans le fichier les valeurs de fluorescence non compensées et la matrice de compensation associée à ces données. La compensation pourra alors être appliquée après l'enregistrement des données.

2.1.3. Annotations

Un autre point important à considérer en amont de l'expérimentation est l'association d'informations expérimentales aux fichiers de cytométrie. Les standards de cytométrie ne couvrent pas cet aspect et ce sera à l'expérimentateur de mettre en place des outils d'annotation appropriés. La plupart des logiciels d'acquisition proposent la création de mots clés qui seront enregistrés avec les données de cytométrie de chacun des fichiers et qui seront ainsi accessibles au moment de l'analyse. Le système de mots clés est souvent utilisé pour la description des anticorps/fluorochromes de l'expérience car ces mots clés sont généralement déjà créés et facilement accessibles lors de l'acquisition. Pour les autres informations expérimentales, il est possible de créer une série de mots clés appropriés (code de l'échantillon, date de prélèvement, etc.), dans le but de faciliter ensuite l'identification et le classement des données et des statistiques issues de l'analyse.

2.2. Standards des fichiers de cytométrie

La Société internationale pour l'avancement de la cytométrie (ISAC, *International Society for Advancement of Cytometry*) a publié son premier standard de cytométrie (FCS, *Flow Cytometry Standard*) en 1984. Ce standard a été mis à jour en 1997 pour donner lieu au FCS 2.0 en 1997, puis le FCS3.0, et plus récemment en 2010 le FCS3.1. Ces mises à jour ont permis de suivre l'évolution de la technologie, en particulier l'augmentation de la taille des fichiers et le développement des réseaux informatiques. Le but de ces standards est de permettre l'échange de données et de favoriser le développement d'outils d'analyse et de traitement des données, notamment par l'apparition de logiciels tiers dédiés à l'analyse. La majorité des cytomètres propose de sauvegarder les données sous un des formats standard. Cependant, même la production d'un format standard n'est pas toujours suffisante pour permettre l'utilisation intuitive et transparente des fichiers de cytométrie. Cela peut s'expliquer par plusieurs facteurs : la coexistence de différents formats, une technologie qui évolue rapidement, des standards peu contraignants, et l'absence de contrôle de la bonne production des standards. Il est donc important de bien vérifier, en amont de l'expérience à réaliser, la compatibilité entre les fichiers produits par le(s) cytomètre(s) sur le(s)quel(s) on fera l'acquisition des données et le logiciel/l'outil d'analyse que l'on souhaite ensuite utiliser.

3. OUTILS POUR L'ANALYSE SUPERVISÉE MANUELLE ET OPTIMISATION DES DONNÉES DE CYTOMÉTRIE

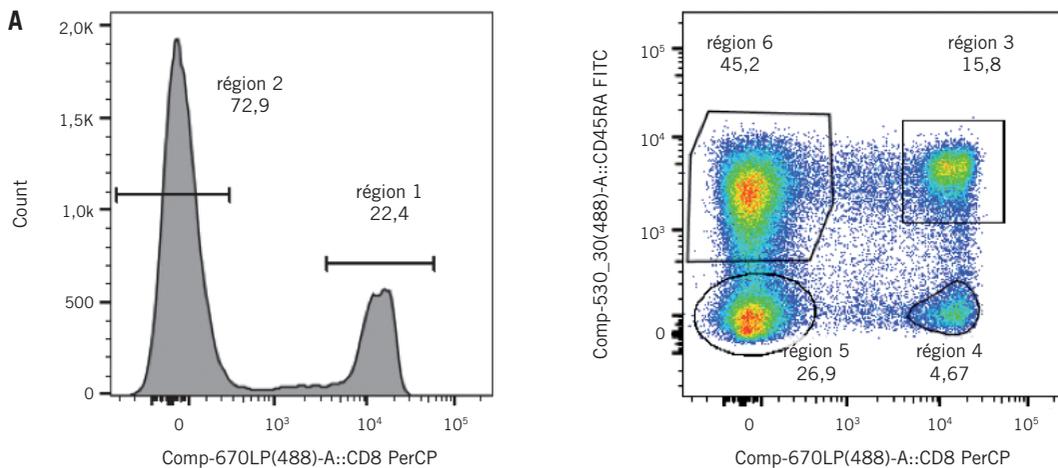
3.1. Bases de la sélection des populations d'intérêt

Les régions d'intérêt (RI) (ou fenêtre, porte) sont utilisées pour isoler une fraction d'événements qui présentent une caractéristique particulière. Une RI permet donc de définir une population qui peut ensuite être analysée plus avant (région d'inclusion). La RI peut également être utilisée pour exclure de l'analyse cette portion d'événement (région d'exclusion). Plusieurs outils graphiques sont à la disposition de l'expérimentateur qui va dessiner ces régions manuellement ou éventuellement à l'aide d'un guide de densité, selon les outils proposés dans le logiciel utilisé. On définit ainsi des populations par l'expression d'un paramètre (histogramme) ou de deux paramètres (graphique bi-varié) (figure 5-5A). En associant les RI, il

est possible de définir des populations complexes caractérisées par l'expression de multiples paramètres. L'association de régions peut se faire par un calcul algébrique à l'aide d'opérateurs booléens (ET, OU, SAUF, en anglais *AND*, *OR*, *NOT*) ou par la création de régions en hiérarchie, une méthode qui met en action l'opérateur « ET » par le simple fait de créer une région à la suite de l'autre (figure 5-5B). La méthode en hiérarchie a l'avantage de définir plus rapidement des populations complexes et de proposer une visualisation structurée des différentes populations créées. Par contre, elle ne permet pas de définir une population qui nécessiterait l'utilisation de l'opérateur booléen « OU ».

À partir des informations enregistrées dans le fichier et des régions ou combinaisons de régions appliquées aux événements de ce fichier, on peut obtenir différentes informations statistiques. Les statistiques extraites de la cytométrie sont liées soit à l'intensité de la fluorescence, soit au

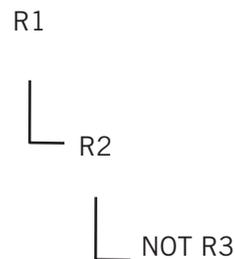
nombre d'évènements collectés. On peut ainsi décrire l'expression d'un marqueur, au sein d'une population définie par une région, en donnant la tendance de cette population (moyenne, médiane...) et la variabilité de cette expression (déviation standard, coefficient de variation...). On peut décrire une population par le nombre d'évènements présents au sein de cette population, ou la fréquence que cette population représente au sein d'une autre population. Il est donc possible par exemple de caractériser les différentes populations par leur nombre absolu ou relatif, ou de comparer le niveau d'expression de marqueurs entre différentes populations. Il est important de garder en tête que l'intensité de fluorescence indiquée dans les fichiers est une *valeur relative* qui ne peut pas être directement corrélée à une quantité absolue de marqueur fluorescent/cible détectée. Seul l'usage de standards de calibration appropriés peut permettre d'obtenir des valeurs absolues.



Définition booléenne de population

R1 AND R2 AND NOT R3

Définition de population en hiérarchie



B

Figure 5-5. Logiques de sélection des populations d'intérêt. (A) Sélection simple de populations sur histogrammes mono ou bi-paramétriques. (B) Combinaison de sélections par utilisation d'opérateurs booléens de type ET, OU, SAUF, ou encore par hiérarchisation.

La cytométrie en flux, technique permettant l'analyse rapide des cellules, a considérablement évolué ces dernières décennies et a contribué à de nombreuses avancées scientifiques et médicales. Ses avancées ont ouvert la porte à des applications devenues incontournables dans différentes spécialités telles que l'hématologie, l'immunologie, la microbiologie, la biologie végétale et marine. La transversalité interdisciplinaire de la cytométrie en flux en fait toute la richesse.

La deuxième édition de **La Cytométrie en Flux** fait le point, en 29 chapitres, sur toutes les nouveautés dues aux avancées instrumentales, à la miniaturisation des lasers, aux progrès de l'optique, de la fluidique mais aussi de l'informatique, des logiciels embarqués et de l'analyse de données, sans oublier les réactifs couvrant maintenant une gamme inimaginable il y a seulement encore quelques années. De nombreux thèmes sont ainsi abordés :

- Principes, techniques, statistiques appliquées à la cytométrie
- Applications dans différents domaines (hématologie, immunologie, détection des microparticules, étude des cellules souches, pharmacologie anticancéreuse...)
- Criblage à haut débit (*high throughput screening*), applications du *fluorescent cell barcoding*
- Assurance qualité, accréditation
- Avancées de la cytométrie couplée à d'autres techniques comme l'analyse spectrale et la spectrométrie de masse
- Gestion des plateformes
- Hygiène et sécurité

Les nombreuses illustrations en couleurs et les protocoles d'analyse viennent compléter ce livre didactique.

Cet ouvrage est destiné aux enseignants et étudiants, chercheurs, biologistes, cliniciens, ingénieurs, techniciens et industriels désireux d'approfondir ou de mettre à jour leurs connaissances.

Il permettra surtout aux utilisateurs non experts d'acquérir les concepts de base de la cytométrie en flux.

Coordonné par **Jean-François Mayol** (AMAL Therapeutics, Genève), **Corinne Laplace-Builhé** (Institut Gustave Roussy, Villejuif) et **Christian D. Muller** (Université de Strasbourg, IPHC-UMR 7178 CNRS, Illkirch), le livre réunit près de 70 auteurs spécialistes de la cytométrie en flux, scientifiques, chercheurs ou industriels.

